

## Firm Size and Growth Rate Variance: The Effects of Data Truncation

Marco Capasso · Elena Cefis

© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** This paper discusses the effects of the existence of natural and/or exogenously imposed thresholds in firm size distributions on estimations of the relation between firm size and the variance in firm growth rates. We argue that these estimations are upwardly biased whenever the threshold operates on the same proxy that is used to calculate the growth rates. We show the potential impact of the bias on simulated data, suggest a methodology to improve these estimations, and present an empirical analysis on Dutch firms. The only stable relation that emerges is the negative relationship between firm size and growth rate variance.

**Keywords** Firm growth · Growth rate variance · Thresholds · Truncation

**JEL Classification** L25 · C21

### 1 Introduction

Several studies on industrial dynamics and firm growth in particular, highlight that the variance in growth rates decreases with increasing firm size. A negative relation between firm size and growth rate variance contradicts the predictions of the Law of

---

M. Capasso  
School of Business and Economics and UNU-MERIT, Maastricht University,  
P.O. Box 616, 6200 MD Maastricht, The Netherlands  
e-mail: marco.capasso@gmail.com

E. Cefis (✉)  
Department of Economics, University of Bergamo, Via dei Caniana 2, 24127 Bergamo, Italy  
e-mail: elena.cefis@unibg.it

E. Cefis  
Urban and Regional Research Centre Utrecht (URU), Utrecht University, P.O. Box 80.115,  
3508 TC Utrecht, The Netherlands

Published online: 31 March 2012

 Springer

Proportionate Effects that the size of a firm and its growth rate are independent. [Meyer and Kuh \(1957\)](#) were the first to observe this negative relation, which was confirmed by other researchers using different datasets over different time periods: for example [Hymer and Pashigian \(1962\)](#), [Stanley et al. \(1996\)](#), [Amaral et al. \(1997, 1998\)](#), [Bottazzi and Secchi \(2003, 2006\)](#), and [Matia et al. \(2004\)](#).

Some recent works, including [Bottazzi et al. \(2002, 2007\)](#), and [Perline et al. \(2006\)](#) find either no relation or find a positive relation between firm size and growth rate variance. [Perline et al. \(2006, p. 8\)](#) observe that, beyond the possible economic explanations, a simple statistical phenomenon emerges when medium or small firms are included in the sample: “The high concentration of small establishments [...] highlights the issue of establishments (or corporations in other studies) that exit from a longitudinal database because they drop to size 0”. This intuition is the starting point of our work.

In our study, we highlight the problems that arise when estimating the relation between size and growth rate variance on panel databases of firms. In particular, we show that an upward bias in the estimation arises whenever, on the same proxy that is used to calculate the growth rates, a threshold operates that restricts the observational range of the data. Most of the studies in which this condition holds, and thus where the bias occurs, are studies in which the number of employees is used both as a proxy for firm size and for calculating the firm growth rates.

The disappearance of small firms in the second year of a growth comparison, either because of the natural lower bound of 0 employees (an ‘endogenous’ threshold, below which firms exit the market), or because of lower bounds that are artificially imposed at higher levels (e.g., an ‘exogenous’ threshold of 20 employees, below which firms are excluded from the data collection and thus cannot be observed), causes an upward bias in the relationship between firm size and the variance of growth rates. We show that the disappearance problem attenuates by running successive regressions between firm size and growth rate variance, where we successively eliminate smaller firms from the starting-year data set (but keep them in the second-year data set). Based on this result, we suggest the use of an alternative methodology that reduces the estimation bias.

The paper is structured as follows. [Section 2](#) describes the model to be estimated. [Section 3](#) illustrates how the bias arises, and provides a numerical example on simulated data. [Section 4](#) describes an alternative methodology to avoid estimation bias, and in [Sect. 5](#) this methodology is applied to a Dutch panel dataset of manufacturing and service firms. [Section 6](#) concludes.

## 2 The Model

In this study, the main variables of interest are the size of the firm and the firm’s rate of growth. For each firm  $j$  and each year  $t$ , we approximate the size  $S_j(t)$  by one plus the number of employees (we add 1 to allow for logarithmic transformations, since self-employment is treated as a firm with 0 employees). Following previous studies, for example [Bottazzi et al. \(2007, 2011\)](#), [Coad \(2007\)](#), and [Coad and Hölzl \(2009\)](#) the rate of growth of firm size  $g_j(t)$  is calculated as the difference in the log size across two consecutive years, namely

$$g_j(t) = \log(S_j(t)) - \log(S_j(t - 1)). \quad (1)$$

The choice of computing growth rates as log size differences is well grounded in the literature on firm growth, since the process describing the evolution of firm size has often been described as a unit root multiplicative process (see for example [Gibrat 1931](#); [Steindl 1965](#); and for a review [Sutton 1997](#)).

The literature commonly tests the relation between firm size and variance in growth rates to divide a sample of firms into several equi-populated size classes. Following [Stanley et al. \(1996\)](#), we model the relation between firm size and variance in growth rates as

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad (2)$$

where  $x_i$  and  $y_i$  respectively are average log size and the log standard deviation of growth rates computed within the size class  $i$ , and  $\varepsilon_i$  is an error term. Size and variance in growth rates are correlated if the slope  $\beta$  is different from zero.

### 3 The Bias

To illustrate the problem that can arise in the presence of endogenous or exogenous thresholds we assume a very simple scenario in which firms can have only two sizes. Suppose our sample contains two equi-populated groups of  $n$  firms. All of the firms that belong to the first group, called  $A$ , have a size equal to  $x_A$ , and all the firms belonging to the second group, called  $B$ , have a size equal to  $x_B$ , with  $x_A < x_B$ .

If we cannot observe firms the size of which is smaller than a given threshold  $\tau < x_A$ , then we are not able to observe (in period 2) all the firms from the first group with a growth rate smaller than  $\log(\tau) - \log(x_A)$  nor all the firms in the second group the growth rate of which is smaller than  $\log(\tau) - \log(x_B)$ . In other words, the observed distribution of growth rates will be left-truncated both for the firms of group  $A$  and for the firms of group  $B$ , and the left-truncation point will be higher for the firms of group  $A$ . As [Baum \(2006, p. 260\)](#) states, “The effect of truncating the distribution of a random variable is clear. The expected value or mean of the truncated random variable moves away from the truncation point, and the variance is reduced”.

Let us assume now that the real distribution of the growth rates is Laplace (double exponential) for both groups  $A$  and  $B$ , with the growth rate variance of group  $A$  not lower than the variance of group  $B$ . In this case, the number of relative frequencies excluded from the observational range is higher for group  $A$  than for group  $B$ . Roughly speaking, the truncated part of the distribution left tail is larger for group  $A$  than for group  $B$ . It is intuitive to say that the negative effect of the truncation on the observed distribution variance is then higher for group  $A$  than for group  $B$ . As a consequence, the difference between the estimated growth rate variances of groups  $B$  and  $A$  is biased upward; and, in turn, the estimated effect of log size on growth rate variance—i.e. the estimation of the  $\beta$  parameter in (2)—is biased upward. The following simulation can show that, for a realistic range of values assumed for  $\alpha$  and  $\beta$ , the bias is far from negligible.

**Table 1** Estimated values  $\hat{\beta}$  of the scaling parameter  $\beta$  linking log size and log standard deviation of growth rates

$\alpha$	$\beta$				
	-0.19	-0.18	-0.17	-0.16	-0.15
0.7	-0.03105	-0.01369	0.00215	0.01795	0.03485
0.8	-0.00522	0.01025	0.02397	0.03851	0.05193
0.9	0.01310	0.02587	0.03810	0.04948	0.06072
1.0	0.02351	0.03476	0.04542	0.05549	0.06524
1.1	0.03117	0.03961	0.04867	0.05787	0.06375

Horizontal axis: true value of the scaling parameter used for the simulation. Vertical axis: value of the intercept of the simulated logsize-logvariance relation

We choose  $n_A = n_B = 10,000$ ;  $S_A = 40$ ;  $S_B = 80$ . We assume that the relation described in the Eq. (2) holds, with  $\alpha$  varying between 0.7 and 1.1, and  $\beta$  varying between  $-0.19$  and  $-0.15$ . The ranges of values for  $\alpha$  and  $\beta$  have been chosen following the empirical findings of Stanley et al. (1996) and of Amaral et al. (1997), who estimate a value of  $\alpha$  equal respectively to 0.8 and 1, and a value of  $\beta$  equal respectively to  $-0.16$  and  $-0.18$ , when using number of employees as a proxy for firm size.<sup>1</sup> Therefore, at each of the 1,000 replications, the growth rates of the 10,000 firms of group  $A$  are drawn from a Laplace distribution having mean equal to zero and log variance equal to  $\alpha + \beta \log(40)$ , and the growth rates of the 10,000 firms of group  $B$  are drawn from a Laplace distribution having mean equal to zero and log variance equal to  $\alpha + \beta \log(80)$ .<sup>2</sup>

In our simulation, we see the effect of a threshold of twenty employees that is imposed on the data observation; that is, we assume that a firm that has less than  $\tau = 20$  employees after the growth event drops out of the sample and cannot be observed after the growth event. If we thus estimate Eq. (2) only on the firms that stay in the observational range, we obtain the estimations  $\hat{\beta}$  shown in Table 1 (mean over the 1,000 replications): The estimated value of  $\beta$  is always higher than the real value. In most of the cases, the upward bias is so strong that a positive relation between log size and log variance of growth rates is detected, instead of the ‘real’ negative relation. Only in a few cases, corresponding to very low values of the ‘real’ scaling parameter  $\beta$  and of the intercept  $\alpha$ , the estimated value of the relation between log size and log variance is negative, but still strongly biased upward.

#### 4 Alternative Methodology

In empirical terms, the truncation can derive from reality (firms cannot have less than zero employees) or from exogenous thresholds deriving from the construction of the

<sup>1</sup> Notice that their data were collected on publicly traded manufacturing firms, so that only the natural boundary of zero firm employees was present, but no additional boundary on firm size was deriving from the data collection.

<sup>2</sup> Simulations were performed using the Matlab software package. The program code is available on request.

database (e.g. if the database collects data only on firms with at least 10 or 20 employees, which is the case for the Community Innovation Survey [CIS] for all European countries). To limit the biases deriving from truncation, we suggest the exclusion from the dataset of firms that are below a given size threshold  $M$  (slightly higher than the threshold already imposed on the data) at time  $t - 1$ , and that the remaining firms are used to construct the balanced sample of growth rates between time  $t - 1$  and time  $t$ . Notice that our artificial threshold is applied only at time  $t - 1$ : no constraint on firm size should be imposed at time  $t$ .

The resulting distribution of growth rates is not bounded from above because, in theory, a firm could grow indefinitely and still belong to the sample, and is not strictly bounded from below because the firm would have to experience a very low negative growth rate (high in absolute value) in order to approach the natural (or endogenous) threshold of zero employees or any exogenous threshold that is imposed by the database construction. This methodology extends the proposal in [Perline et al. \(2006\)](#) simply to exclude from the regression the first size class (i.e. associated with the smallest average size). In order to show how our methodology applies with respect to real data, and to calibrate  $M$  for the reduction of the number of firms that are excluded from the dataset, we conduct the following empirical analysis.

## 5 The Empirical Analysis

### 5.1 The Case of an Endogenous Threshold

The data in this paper were collected by the Centraal Bureau voor de Statistiek (CBS) and stem from the Business Register of enterprises. The Business Register database includes all firms registered for fiscal purposes in the Netherlands in the year considered. The population includes firms with zero employees, referred to as self-employment. We consider growth rates between 2002 and 2003 for all Dutch firms in manufacturing and services (approximately 60,000 manufacturing and 1,000,000 services firms), considered separately: that is, the two groups belong to two different macro-sectors. Therefore, the data do not have any exogenous lower threshold but just the endogenous natural limit of zero employees.

In a first step, we exclude from the sample all firms with less than  $M$  employees at time  $t - 1$  (i.e., in year 2002) and retain all of the remaining firms that still exist at time  $t$  (i.e., in 2003).  $M$  is an integer value varying between 0 and 20, and we consider all of the cases between 0 and 20: When  $M$  is equal to 0, no firms are excluded by the artificial threshold, and we retain all firms that exist in 2002 and 2003. In a second step we estimate the relation between firm size and growth rate variance (Eq. 2) by minimizing the Least Absolute Deviation of the error terms  $\varepsilon_i$ , assuming that the residuals are Laplace distributed as in [Bottazzi et al. \(2007\)](#).<sup>3</sup> 25 equi-populated size classes will be used for estimating (2).

---

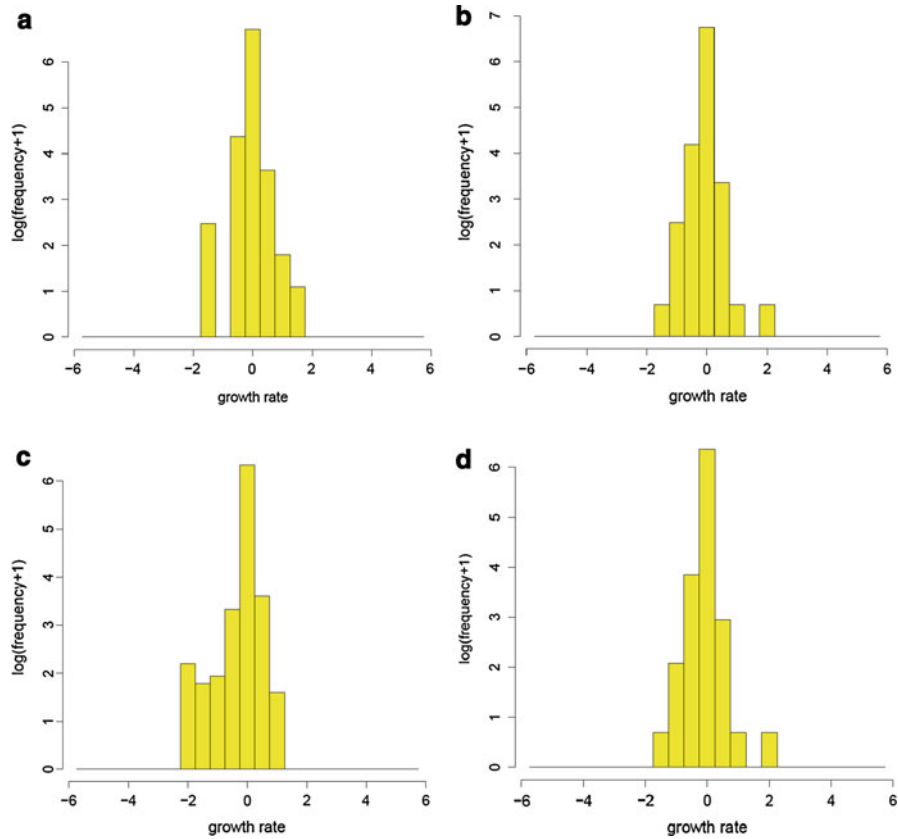
<sup>3</sup> The variables that are used in the regressions were built using the R software package. The LAD regressions were run using the *qreg* function of the Stata software package.

**Table 2** Empirical relation between size and growth rate variance: The case of an endogenous threshold (whole population)

Threshold $M$	Manufacturing				Services			
	Coefficient	SE	$t$ ratio	$p$ value	Coefficient	SE	$t$ ratio	$p$ value
0	0.2307	0.0726	3.18	0.004	0.2062	0.0685	3.01	0.006
1	0.2640	0.0556	4.75	0.000	0.4570	0.0234	19.54	0.000
2	0.1715	0.0429	4.00	0.001	0.3309	0.0696	4.76	0.000
3	0.0394	0.0658	0.60	0.555	0.2090	0.0695	3.01	0.006
4	-0.0201	0.0546	-0.37	0.715	0.0999	0.0581	1.72	0.099
5	-0.0200	0.0391	-0.51	0.614	0.0807	0.0453	1.78	0.088
6	-0.0979	0.0718	-1.36	0.186	0.0691	0.0378	1.83	0.081
7	-0.1411	0.0603	-2.34	0.028	0.0010	0.0635	0.02	0.987
8	-0.1561	0.0730	-2.14	0.043	-0.0247	0.0403	-0.61	0.546
9	-0.2203	0.0491	-4.49	0.000	-0.1058	0.0253	-4.18	0.000
10	-0.2025	0.0344	-5.89	0.000	-0.1141	0.0387	-2.95	0.007
11	-0.2401	0.0493	-4.87	0.000	-0.1063	0.0154	-6.89	0.000
12	-0.2635	0.0494	-5.33	0.000	-0.1112	0.0410	-2.71	0.012
13	-0.2077	0.0474	-4.38	0.000	-0.1284	0.0208	-6.17	0.000
14	-0.2165	0.0640	-3.38	0.003	-0.1126	0.0458	-2.46	0.022
15	-0.2920	0.0523	-5.58	0.000	-0.1472	0.0276	-5.34	0.000
16	-0.2045	0.0454	-4.51	0.000	-0.0638	0.0419	-1.52	0.142
17	-0.2138	0.0500	-4.28	0.000	-0.0471	0.0378	-1.25	0.225
18	-0.2700	0.0786	-3.43	0.002	-0.0199	0.0393	-0.51	0.618
19	-0.1955	0.0627	-3.12	0.005	-0.0833	0.0258	-3.23	0.004
20	-0.2155	0.0848	-2.54	0.018	-0.0603	0.0325	-1.86	0.076

Table 2 shows that the estimated coefficient  $\beta$  seems to decrease with  $M$ , with a clear tendency to move from positive to negative values as the threshold increases. Our methodology consists of excluding from the database the firms that have employees that are, at time  $t - 1$ , below a numerical threshold  $M$ , letting  $M$  increase until the estimated coefficient is stable (i.e., keeping the smallest  $M$  for which the estimated  $\beta$  finds a plateau). It should be possible to achieve stability for a reasonably small  $M$ ; in our case  $M = 9$  for the manufacturing and services sectors. If estimation of the coefficient  $\beta$  does not stabilize at small values of  $M$ , model (3) is probably mis-specified (i.e., there are nonlinearities in the relation being studied). Our methodology allows us to observe that the relationship between firm size and variance in firm's growth rates is stable only for negative values of the coefficient.<sup>4</sup> A positive relation is due

<sup>4</sup> An additional analysis has been conducted at a disaggregated level, in particular on two 2-digit sectors (SBI sectors 15 and 50, where SBI—i.e., Standard Bedrijfsindeling—is the Dutch correspondent of SIC codes). In both cases the 'artificial' threshold  $M$  has the same effect as in the general analysis on manufacturing and services: as  $M$  increases, first the sign of the estimated relation between size and growth rate variances changes from positive to negative, and then the value of the estimated parameter becomes more stable.



**Fig. 1** Histograms of the growth rate distributions for the first and the last size classes of manufacturing firms, when  $M = 3$  and  $M = 6$  (imposed thresholds on the initial number of employees). **a** First size class,  $M = 3$ . **b** Last size class,  $M = 3$ . **c** First size class,  $M = 6$ . **d** Last size class,  $M = 6$

only to the effects of truncation, which flow from the existence of endogenous and/or exogenous thresholds that bias the estimation upwards.

Figure 1 helps to explain how our methodology corrects the bias, in practical terms. We focus on the population of manufacturing firms, and consider the two cases of  $M = 3$  and  $M = 6$ . For each of the two cases, we consider only the firms that have employees that are equal to or higher in number than  $M$ , and we divide them into 25 equi-populated size classes (in increasing order of average size), as we would normally do before estimating model (2). Figure 1a, b show the distribution of growth rates of firms belonging respectively to the first and last size classes (i.e., the classes with smallest and largest average size) for the case of  $M = 3$ . Figure 1c, d show the distribution of growth rates of firms belonging respectively to the first and last size for the case of  $M = 6$ .

It is evident that, for the size class corresponding to the highest average firm size, the distribution of growth rates is similar for  $M = 3$  and  $M = 6$  (Fig. 1b, d). This is due to the fact that only a few firms having an initial high number of employees

suffer a size decrease that is so dramatic that the firm exits the observational range that is defined by the natural threshold of zero employees. In other words, the observed distribution of growth rates of bigger firms is not strongly influenced by any truncation effect that is caused by firm exit, and therefore our imposed threshold defined by  $M$  is not correcting any truncation effect: the observed growth rate variance of bigger firms is not influenced by the ‘artificial’ threshold  $M$ .

For the smaller firms the picture is completely different. Figure 1a, c show that the distribution of the smaller firms’ growth rates is strongly affected by the truncation caused by firm exit, leading to a strong decrease of variance. However, the effect of truncation is much stronger when the  $M$  threshold imposed by us is lower (Fig. 1a, where  $M = 3$ ) than when the threshold  $M$  is higher (Fig. 1c, where  $M = 6$ ). In other words, a higher  $M$  allows us to reduce the effect of dropouts on the left-truncation of the growth rate distribution, and thus to reduce the downward influence of dropouts on the observed growth rate variance, for the size classes that are associated with low firm sizes.

The comparison between Fig. 1a, b shows that the increase in growth rate variance associated with an increase in firm size, observed when  $M = 3$ , is mainly due to the fact that many small firms cannot experience very low growth rates without exiting the market (or, in general, the observational range). That is: for  $M = 3$ , our methodology is still not able to correct for the dropout bias, because the left-truncation of the growth rate distribution is still affecting strongly the observed growth rate variance of the first size class. For  $M = 6$ , the left-truncation of the growth rate distribution of the first size class (Fig. 1c) does not have a dramatic effect on the observed growth rate variance. Indeed, the statistics of Table 3 (left panel corresponding to Fig. 1a, b, right panel corresponding to Fig. 1c, d) suggest that the variance of the first size class is higher for  $M = 6$  than for  $M = 3$  because the first ten percentiles of the growth rate distribution are shifted downward, which in turn depends on the fact that the left-truncation of the distribution occurs at a lower value of the growth rate. The downward shift of the first ten percentiles increases the growth rate variance of the first class, and the upward bias in the relation between firm size and growth rate variance is then reduced.

## 5.2 The Case of an Exogenous Threshold

We now turn to the case of an exogenous threshold, occurring when data have been collected only for firms that have employees that are greater in number than a given lower bound (as for the cases of databases that do not include micro and small firms). In other words, firms that have less than a given number of employees at time  $t - 1$ , or at time  $t$ , or both, are excluded from the 2-year panel and thus cannot be observed.

In this case, we must apply a slight modification to our procedure described in the previous subsection: Instead of approximating firm size simply by adding one to the number of employees, as was previously done in order to allow logarithmic transformations for self-employed firms (having zero employees), we approximate firm size by one plus the number of employees minus the value of the exogenous threshold. Roughly speaking, we measure firm size as a distance from the exogenous threshold rather than as a distance from the endogenous threshold of zero employees. Apart



**Table 3** Statistics of the growth rate distributions for the first and the last size classes of manufacturing firms, when  $M = 3$  and  $M = 6$  (imposed thresholds on the initial number of employees)

	$M = 3$		$M = 6$	
	First size class	Last size class	First size class	Last size class
Number of firms	954	954	653	653
Mean	-0.0127	-0.0387	-0.0389	-0.0361
Variance	0.0600	0.0387	0.0855	0.0405
Skewness	-1.3804	-0.5282	-3.8751	0.0699
Kurtosis	18.4024	22.5843	26.0127	25.3170
Minimum	-1.3863	-1.3298	-1.9459	-1.3298
10th percentile	0	-0.2072	-0.1542	-0.2163
20th percentile	0	-0.1022	0	-0.0993
30th percentile	0	-0.0391	0	-0.0369
40th percentile	0	-0.0066	0	-0.0064
50th percentile	0	0	0	0
60th percentile	0	0	0	0
70th percentile	0	0	0	0
80th percentile	0	0.0126	0	0.0188
90th percentile	0	0.0863	0	0.0911
Maximum	1.6582	1.9674	0.9985	1.9674

from this particular way of measuring firm size, our methodology is exactly the same as for the case of the endogenous threshold, described in the previous subsection. We can explain the reason behind such technical device by a simple example.

Let us come back, first, to the case of endogenous threshold, and consider the example of a firm  $j$  having 10 employees at time  $t - 1$ . At time  $t$ , this firm can have a minimum number of employees equal to zero (in case it becomes self-employed). If we simply approximate the firm size by adding one to the number of employees (for allowing logarithmic transformations), the firm of our example can have a minimum growth rate of  $\log(S_j(t)) - \log(S_j(t-1)) = \log(1+0) - \log(1+10) = -2.40$ . If this firm were the largest firm of the first size class at time  $t - 1$ , then the left-truncation of the growth rate distribution, for the first size class, would be at  $-2.40$ .

Suppose now that the data are bounded by an exogenous lower threshold of five employees, i.e. only firms having at least five employees (both at time  $t - 1$  and at time  $t$ ) can be observed. In this case, the firm of our example will be included in the database only when its minimum growth rate is  $\log(S_j(t)) - \log(S_j(t-1)) = \log(1+5) - \log(1+10) = -0.61$ . If this firm were the largest firm of the first size class at time  $t - 1$ , then the left-truncation of the (observed) growth rate distribution for the first size class would be at  $-0.61$ —i.e., at a value that is much higher than for the case of an endogenous threshold—thus leading to a very strong downward bias in the variance estimation for the growth rate distribution of the first size class, and in turn to a strong upward bias in the estimation of the relation between firm size and growth rate variance.

If we had to apply directly the procedure described in the previous subsection, our artificial threshold  $M$  should be very high in order to erase the effect of the left-truncation in the growth rate distribution. In particular, in order to reach the previous left-truncation value of  $-2.40$ , the largest firm of the first size class should have 65 employees at time  $t - 1$  as  $-2.40 = \log(1) - \log(11) = \log(6) - \log(66)$ , i.e. our artificial threshold  $M$  should be high enough to exclude all of the firms that have less than 65 employees at time  $t - 1$  from the analysis.

Our point is the following: The customary use of log differences to proxy firm growth rates creates a hidden link between the (endogenous or exogenous) data thresholds and the left-truncation of the growth rate distributions. This link acts through the first term of the log-difference, and creates a much higher estimation bias when the data lower bound is higher, which is the case for exogenous thresholds in the data collection.

Fortunately, we can easily reduce the extent of the problem by approximating firm size by one plus the number of employees minus the value of the exogenous threshold. For example, if the data allow us to observe only firms having at least five employees at time  $t - 1$  and at time  $t$ , then a firm having 10 employees at time  $t - 1$  has a size, according to our new definition, equal to  $(1 + 10 - 5)$ , and can reach a minimum observable size, at time  $t$ , of  $(1 + 5 - 5)$ . Its minimum growth rate, in order not to be excluded from the data collection, would be  $\log(1 + 5 - 5) - \log(1 + 10 - 5) = -1.79$ , and so the left tail of the growth rate distribution would be truncated at  $-1.79$  rather than  $-0.61$ .

Therefore, by measuring firm size as a distance from the exogenous threshold, we need a much lower value of the artificial threshold  $M$ , and a much lower number of excluded firms, to reduce the upward bias in the estimation. Indeed, this way of measuring firm size reduces the extent of the left-truncation problem back to the case of the endogenous threshold, so that the methodology described in Sect. 4 is able to work well without further modifications, and with the exclusion of a relatively low number of firms.

To show the effectiveness of this method in practical terms, we use the same data as in the previous subsection, but we now apply an exogenous threshold of five employees. In other words, we exclude from the panel the firms that have less than five employees at time  $t - 1$ , or at time  $t$ , or at both  $t - 1$  and  $t$ , in order to simulate a lower bound of five employees exogenously imposed in the data collection. Following the intuition of the previous paragraph, we approximate firm size by one plus the firm's number of employees minus five (the exogenous threshold), and we then apply our methodology as described in Sect. 4.

The results are in Table 4. Manufacturing firms exhibit a strong negative relation between size and growth rate variance, even when the artificial threshold  $M$  is not excluding any additional firm (i.e., when  $M$  is equal to five, the same value of the exogenous threshold). Coefficients are smaller, in absolute value, than was observed in the previous subsection, but they are still negative and significant. For the case of services, we obtain significant, and negative, coefficients for values of  $M$  that are equal to or higher than nine. This corresponds to what was observed in the previous subsection (without an exogenous threshold), although now the negative significant coefficients are lower in absolute value. In other words, our methodology still detects

**Table 4** Empirical relation between size and growth rate variance: The case of an exogenous threshold (considering only firms that have at least five employees in both years 2002 and 2003)

Threshold $M$	Manufacturing				Services			
	Coefficient	SE	$t$ ratio	$p$ value	Coefficient	SE	$t$ ratio	$p$ value
5	-0.1005	0.0273	-3.68	0.001	-0.0107	0.0208	-0.51	0.612
6	-0.0800	0.0257	-3.11	0.005	-0.0144	0.0168	-0.86	0.400
7	-0.0745	0.0320	-2.33	0.029	-0.0074	0.0251	-0.30	0.771
8	-0.1034	0.0311	-3.33	0.003	-0.0114	0.0188	-0.61	0.551
9	-0.1077	0.0237	-4.53	0.000	-0.0504	0.0198	-2.55	0.018
10	-0.1242	0.0314	-3.96	0.001	-0.0660	0.0198	-3.32	0.003
11	-0.1086	0.0261	-4.15	0.000	-0.0773	0.0140	-5.54	0.000
12	-0.1089	0.0312	-3.49	0.002	-0.0433	0.0233	-1.86	0.076
13	-0.0987	0.0289	-3.42	0.002	-0.0466	0.0199	-2.35	0.028
14	-0.1097	0.0410	-2.68	0.013	-0.0581	0.0164	-3.55	0.002
15	-0.1513	0.0520	-2.91	0.008	-0.0579	0.0240	-2.42	0.024
16	-0.1263	0.0242	-5.21	0.000	-0.0397	0.0244	-1.62	0.118
17	-0.1562	0.0443	-3.52	0.002	-0.0341	0.0183	-1.87	0.075
18	-0.1491	0.0637	-2.34	0.028	-0.0340	0.0257	-1.33	0.198
19	-0.1541	0.0613	-2.51	0.019	-0.0272	0.0303	-0.90	0.378
20	-0.1285	0.0396	-3.24	0.004	-0.0301	0.0268	-1.13	0.272

a negative relation between firm size and growth rate variance even after replicating the exogenous lower bounds that some data collection services impose, and without increasing the number of firms that are excluded from the analysis.

## 6 Conclusions

We have shown that a bias can arise in the estimation of the relation between firm size and variance in growth rates, when endogenous and/or exogenous thresholds truncate the firm size distribution; that is, when micro firms are included in the analysis or when the dataset considers only firms with sizes that are above a certain threshold in terms of numbers of employees. In particular, we show that an upward bias in the estimation exists whenever the threshold operates on the proxy that is used to construct the growth rates. This problem was highlighted by [Perline et al. \(2006\)](#), but is ignored in most of the literature. However, it seems to be a determinant of the explanation of the contrasting evidence from previous studies.

After pointing to the bias that can follow from a truncated firm's size distribution, from a theoretical perspective, we suggested a simple methodology to estimate the relation between firm size and the variance in firm growth rates that avoids the bias. We tested our methodology on the population of manufacturing and service firms in the Netherlands. The results show that the methodology that we propose allows us to observe how the estimated coefficients of the size-growth rate variance relation change

depending on the firm size threshold. We suggest using an artificial threshold  $M$  that can increase until the estimated coefficient is stable and retaining the lowest  $M$  for which the estimated  $\beta$  finds a plateau.

Our empirical analysis shows that there is a stable, negative relationship between firm size and variance in growth rates: The dynamics of smaller firms are characterized by a stronger turbulence. Many recent studies show a positive relationship, but this is due to the presence of natural/endogenous or exogenous thresholds in the firm size distributions.

**Acknowledgments** The authors thank Lawrence J. White (general editor), two anonymous referees, Andrew Bernard, Koen Frenken, and the participants at the DIME Workshop on “The dynamics of firm evolution: productivity, profitability and growth”, Pisa, 3–4 October 2008. The empirical work was carried out at the Centre for Research of Economic Microdata at Statistics Netherlands (CBS). The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands. The authors thank Gerhard Meinen and on-site staff of CBS for their collaboration. This work was supported by Utrecht University [High Potential Grant (HIPO) to E. Cefis and K. Frenken]; and the University of Bergamo [grant ex 60 %, n. 60CEF11, Department of Economics, to E. Cefis].

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Salinger, M. A., & Stanley, H. E. (1998). Power law scaling for a system of interacting units with complex internal structure. *Physical Review Letters*, *80*(7), 1385–1388.
- Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Salinger, M. A., Stanley, H. E., & Stanley, M. H. R. (1997). Scaling behavior in economics: The problem of quantifying company growth. *Physica A*, *244*, 1–24.
- Baum, C. F. (2006). *An introduction to modern econometrics using Stata*. College station, TX: Stata Press.
- Bottazzi, G., Cefis, E., & Dosi, G. (2002). Corporate growth and industrial structure: Some evidence from the Italian manufacturing industry. *Industrial and Corporate Change*, *11*, 705–723.
- Bottazzi, G., Cefis, E., Dosi, G., & Secchi, A. (2007). Invariances and diversities in the evolution of manufacturing industries. *Small Business Economics*, *29*, 137–159.
- Bottazzi, G., Coad, A., Jacoby, N., & Secchi, A. (2011). Corporate growth and industrial dynamics: Evidence from French manufacturing. *Applied Economics*, *43*(1), 103–116.
- Bottazzi, G., & Secchi, A. (2003). Common properties and sectoral specificities in the dynamics of U.S. manufacturing companies. *Review of Industrial Organization*, *23*, 217–232.
- Bottazzi, G., & Secchi, A. (2006). Explaining the distribution of firm growth rates. *Rand Journal of Economics*, *37*(2), 235–256.
- Coad, A. (2007). A closer look at serial growth rate correlation. *Review of Industrial Organization*, *31*(1), 69–82.
- Coad, A., & Hözl, W. (2009). On the autocorrelation of growth rates. *Journal of Industry, Competition and Trade*, *9*(2), 139–166.
- Gibrat, R. (1931). *Les inégalités économiques*. Paris: Librairie du Recueil Sirey.
- Hyer, S., & Pashigian, P. (1962). Firm size and the rate of growth. *Journal of Political Economy*, *70*(4), 556–569.
- Matia, K., Fu, D., Buldyrev, S. V., Pammolli, F., Riccaboni, M., & Stanley, H. E. (2004). Statistical properties of business firms structure and growth. *Europhysics Letters*, *67*(3), 498–503.
- Meyer, J. R., & Kuh, E. (1957). *The investment decision*. Cambridge, Massachusetts: Harvard University Press.

- Perline, R., Axtell, R., & Teitelbaum, D. (2006). *Volatility and asymmetry of small firm growth rates over increasing time frames*. U.S. Small Business Administration, The Office of Advocacy Small Business Working Papers 06/rarpdt. Retrieved October 1, 2008, from <http://ideas.repec.org/p/sba/wpaper/06rarpdt.html>.
- Stanley, M. H. R., Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Leschorn, H., Maass, P., Salinger, M. A., & Stanley, H. E. (1996). Scaling behavior in the growth of companies. *Nature*, 379, 804–806.
- Steindl, J. (1965). *Random processes and the growth of firms: A study of the Pareto law*. London: Griffin.
- Sutton, J. (1997). Gibrat's Legacy. *Journal of Economic Literature*, 35, 40–59.